

Copyright  
by  
Qiqian Yang  
2014

The Thesis committee for Qiqian Yang

Certifies that this is the approved version of the following thesis:

## A Station-Level Analysis of Rail Transit Ridership in Austin

APPROVED BY

SUPERVISING COMMITTEE:

Supervisor: \_\_\_\_\_

Junfeng Jiao

Co-supervisor: \_\_\_\_\_

Daniel Yang

# **A Station-Level Analysis of Rail Transit Ridership in Austin**

**by**

**Qiqian Yang, B.E.C.E.**

**Thesis**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**Master of Science in Community and Regional Planning**

**The University of Texas at Austin**

**August 2014**

## **Acknowledgments**

I would like to thank my advisor Dr. Junfeng Jiao.

Thank you my fellows at Capital Area Metropolitan Planning Organization, notably my supervisor Dr. Daniel Yang. Without his help, I won't be able to finish this thesis.

I would like to thank Todd Hemingson and Jeniffer L. Govea from Capital Metro as well as my fellows at Center for Transportation Research. I would like to thank you my dear friends for providing technical assistants.

I would like to express my appreciations to Community and Regional Planning department at Architecture School for accepting me as a master student.

Last but not least, thank you my parents for supporting me to accomplish two years' study at UT Austin. Without their encouragement and selfless support, I won't have the chance to study in U.S.

## **Abstract**

### **A Station-Level Analysis of Rail Transit Ridership in Austin**

Qiqian Yang, M.S.C.R.P.

University of Texas at Austin, 2014

Supervisor: Junfeng Jiao

In the past two decades, Austin has tremendous population growth, job opportunity in the downtown core and transportation challenges associated with that. Public transit, and particularly rail, often is regarded as a strategy to help reduce urban traffic congestion. The Urban Rail, which combines features of streetcars and light rail, is introduced into Austin as a new transit rail. The City of Austin, Capital Metro and Lone Star Rail are actively studying routing, financial, environmental and community elements associated with a first phase of Urban Rail.

This thesis collected 2010 Origin and Destination Rail Transit Survey data from Capital Metropolitan Transportation Authority. The research focuses on the rail transit ridership. Two regression models are applied to analyze the factors influencing Austin rail transit ridership. One model is focusing on the socioeconomic characteristics. One model is focusing on the spatial factors.

Our model shows that demographic factors have more significant effect than spatial factors.

In addition, this work also tries to analyze the correlations between those factors and make recommendations based on the analysis result.

# Contents

Chapter One: Introduction .....	1
1.1 Why Austin .....	1
1.2 Sections of this thesis.....	3
Chapter Two: Research Questions .....	5
Chapter Three: Literature review.....	6
3.1 Research conduct method .....	6
3.2 Modeling method.....	6
Direct Ridership Model .....	6
3.3 Factors influencing the ridership .....	8
Chapter Four: Data .....	11
4.1 Where did the data come from?.....	11
4.2 Data description .....	11
4.3 Data cleaning/data categorization.....	12
Chapter Five: Modeling and Analysis .....	20
5.1 Method of modeling .....	20
5.2 Multiple Linear regression model.....	22
Demographic model.....	23
Spatial model .....	29
5.3 Correlation Table Analysis.....	35
Chapter Six: Conclusion .....	39
Chapter Seven: Limitation and future work .....	40
7.1 Data limitation: .....	40
7.2 Model limitation: .....	40
7.3 Observation limitation .....	40

7.4 Future work.....	41
Appendix.....	42
Appendix A: Questionnaire .....	42
Appendix B: Question Name Definitions.....	47
Bibliography .....	49

## List of Tables

Table 4.3.1 Spatial options definition .....	17
Table 4.3.2 Demographic characteristic options definition .....	18
Table 5.2.1 Demographic model output .....	25
Table 5.2.2 Spatial model output .....	30
Table 5.3.1 Correlation table of demographic model.....	37
Table 5.3.2 Correlation table of spatial model .....	38



## List of Figures

Figure 1.1.1 Project Connect Proposed Map .....	3
Figure 4.3.1 Walking distance (1).....	12
Figure 4.3.2 Walking distance (2).....	12
Figure 4.3.3 Traveling distance (1).....	13
Figure 4.3.4 Traveling distance (2).....	13
Figure 4.3.5 Age .....	14
Figure 4.3.6 Categorized age option.....	14
Figure 5.2.1 Car ownership .....	27
Figure 5.2.2 Income .....	27
Figure 5.2.3 Education land and transit rail in Austin .....	32
Figure 5.2.4 Walking distance (3).....	33
Figure 5.2.5 Walking distance (4).....	35

# **Chapter One: Introduction**

## **1.1 Why Austin**

In the past two decades, Austin has tremendous population growth, job opportunity in the downtown core and transportation challenges associated with that. Central Texas is on a lot of “best of” lists, but traffic isn’t one of them. The 2012-13 INRIX Traffic Scorecard Annual Report (INC. 2013) ranks Austin fourth in the U.S for time wasted in traffic, worse than Houston and New York City.

Public transit, and particularly rail, often is regarded as a strategy to help reduce urban traffic congestion, especially in these times of economic downturn, rising gas prices, pollution, and growing awareness of global climate change.

The Central Corridor—roughly bounded by RM 2222/Koenig Lane, MoPac, Oltorf Street and Springdale Road—is part of the overarching Project Connect regional transit plan that involves a partnership among the city of Austin, Capital Metro and regional group Lone Star Rail District. Project Connect’s goal is to provide a high-capacity transit system that links cities in Central Texas via bus, express lanes and commuter, regional and urban rail. The vision map indicates Austin could have 40 miles of urban rail. Based on Project Connect’s homepage, the Central Corridor has a wealth of factors that contribute to its congestion problems, including a “ring of congestion” around I-35, Lamar Boulevard and MoPac. Other factors that contribute to congestion include 47,000 daily work trips, an average of three festivals per week and a total of 197,000 jobs, which represents 23 percent of the region’s total jobs.

According to Austin Urban Rail homepage, Urban Rail --- combines features of streetcars and light rail— is a way to connect Central Texans to destinations at the

heart of our region (Coorportation, 2013). The University of Texas, the State Capitol Complex and Downtown are all located in central Austin, making it the largest employment center in the region. Operating primarily in its own lane, Urban Rail is expected to provide better mobility to, from and within Austin's urban core, while providing seamless connections between the services and systems that tie together the regional Project Connect vision.

The Figure 1.1.1 shows the Project Connect Proposed Map. (Connect, 2014)



Figure 1.1.1: Project Connect Proposed Map (Connect, 2014)

## 1.2 Sections of this thesis

The development of a successful transit system must include a ridership analysis during the planning stage. So what explains transit ridership? The answer to this

simple question is both obvious and complex. In chapter three, some important previous research works are examined. Then I discuss this thesis's data sources as well as describe data cleaning and variables definition method in chapter four.

In the chapter five, multiple linear regression models were used to analyze the data. Because the sample size is too small, I used the ridge regression method and m5' feature selection method to develop relatively good final models. After analyzing important variables output from the final models, I create correlation tables for both of the models to confirm/reject the possible assumptions

Conclusion, study limitation and future work can be found in the chapters six and seven.

## **Chapter Two: Research Questions**

The main research purpose of this thesis is to identify the factors affecting rail ridership.

This study also examines the correlations between demographic characteristics and spatial environment factors.

This study will explore the correlation between demographic characteristics and spatial environment. This study encompasses the two models to examine the impacts of origin/destination factors on Metro trip patterns. In this study I am not going to discuss the station to station level ridership, but will focus on the station level ridership analysis.

There are two ways to conduct a research analysis of ridership: descriptive analyses and causal analysis. Descriptive analysis method uses the qualitative survey and interview data. Studies based on operators' perceptions tend to emphasize the variables such as: service improvements and adjustments; fare innovation and changes; marketing and information; new planning approaches and partnership; service quality and coordination. (Brain D. Taylor, Camille N.Y. Fink, 2001) Causal analysis method generally analyze with multivariate regressions (multi-variables). Models include combinations of various variables; Causal method, which allow researcher to collect data from large number of agency, could be more sophisticated empirical compared with the descriptive method.

## **Chapter Three: Literature review**

### **3.1 Research conduct method**

There are two ways to conduct a research analysis of ridership: descriptive analyses and causal analysis. Descriptive analysis method uses the qualitative survey and interview data. Studies based on operators' perceptions tend to emphasize the variables such as: service improvements and adjustments; fare innovation and changes; marketing and information; new planning approaches and partnership; service quality and coordination. (Brain D. Taylor, Camille N.Y. Fink, 2001) Causal analysis method generally analyze with multivariate regressions (multi-variables). Models include combinations of various variables; Causal method, which allow researcher to collect data from large number of agency, could be more sophisticated empirical compared with the descriptive method.

### **3.2 Modeling method**

#### **Direct Ridership Model**

The four step model has been the standard procedure for transit forecasting for over 50 years. Four step model is also named trip based model. To form the principle database, model focused on trip origin-destination (O-D) rather than activity surveys. The application of this modeling approach is near universal. In theory, the complexity of these models makes them the best tools for evaluating new transit facilities, but in practice there are several potential problems (Norm Marshall, Brain Grady, 2006) such as model accuracy (in most regional models much more attention has been paid to matching traffic counts on individual roadway segments than on

matching transit loadings on individual route segments or stations), travel input data (estimation is typically based on relatively old household surveys, which may include only a small number of transit trips in the area of interest), sensitivity to land use (regional models are generally insensitive to land use), institutional barriers (transit providers are often not part of the modeling process) and cost of use (four-step travel models are cumbersome and expensive).

In recent decades, researchers have developed ridership forecasting regression models as alternatives to the costly and time consuming four step model. Direct models based on multiple regression analysis are a complementary approach to estimating ridership as a function of station environment and transit services features (Kuby, Barranda, & Upchurch, 2004) (Chu, 2004) (Robert Cervero, 2010). It includes station area demographics, central business district (CBD) employment, and the station areas' built environments to analyze ridership.

Direct demand model based on OLS multiple regression analysis is adopted to evaluate then impacts of chosen variables on ridership at station level. OLS regression can handle both numerical and dummy variables, which are flexible, widely used, and easily understood (Kevin Manaugh, 2010); (Rober Hannay, 2006); (Kuby, Barranda, & Upchurch, 2004);

Such models are a quick-response and less expensive alternative to the four step model. They also capture better the influences on travel demand of the built environment of the station. Traditional regional demand models are sensitive to changes in land use, but their resolution tends to be too coarse to pick up fine-grained design and land-use mix features of neighborhood-scale initiatives such as new urban planning and transit-oriented developments. Direct ridership models based on the use of regression analysis combine features from all four steps of the traditional travel model. It is clearly not as comprehensive and systematic as the four-step process. Yet its transparency and high explanation power may make it useful for experimenting with different planning scenarios. And they are directly



responsive to land-use characteristics within the station catchment areas. (Cervero, Robert, 2006)

Direct ridership models generally have small sample sizes since observations consist of transit stations or stops. Thus degree-of-freedom constraints often limit the number of variables. It is because of these limitations that direct models fall under the rubric of sketch planning tools.

### **3.3 Factors influencing the ridership**

From the point of view of direct estimation models at station level, factors affecting ridership can be classified into two types: socioeconomic factors and spatial factors. The spatial factors are thought to influence travel demand along three principal dimensions: density, diversity, and design (Robert Cervero, Kara Kockleman, 1997). Urban density is the critical driver of transit ridership. The evidence for a positive relationship between population density and transit ridership is well established at station level (Samuel Seskin, Cervero, Robert, 1996). The significance of urban density is that the more people living and/or working in close proximity to transit, the greater the likelihood the service will be used (Alan T. Murraya, Rex Davisb, Robert J. Stimsonc, Luis Ferreirad, 1998). Most of diversity factors focus on land-use type and mix. For example in Brinckerhoff's (Brinckerhoff, 1996) study concluded that land use type do influencing transit use, although less so than density. Land-use mix produces a more balanced demand for public transport, with reducing differences between peak and off-peak periods (Cervero, 2004). Filion (Filion, 2001) found that mixed-use suburban centers have been successful in attaining higher transit use than the typical suburban area. Moreover, neighborhoods that are more walkable favor access to stations on foot and increase transit ridership (Robert Cervero, Michael Duncan, 2008). The walkability and transit station design can be considered as design factors.

Studies have been identifying a host of socioeconomic factors. Not surprisingly, income level and pricing factors are most frequently analyzed in studies of the factors influencing the rail transit ridership. Some of them are focusing on pricing factors. Kain and Liu introduce fare variable in the Portland, San Diego and Houston's case study, and finally find that a combination of factors such as fare, employment and gas price contribute to transit ridership. (Kain, 1999) Similarly, Gomez-Ibanez (Gomez-Ibanez, 1996) finds the ridership between 1970 and 1990 in Boston is not largely influencing by the fare or income factors. Employment was more significant than the per capita income.

Despite income level, car ownership are significant factors frequently included in the regression model. Ordinary, various researches use per capita passenger car registrations or percent carless household as the measures as auto ownership.

Kitamura used a sample from Dutch National Mobility Panel survey to examine the study of car ownership and transit use. The finding suggests that "the increase in car use, which is a consequence of increasing car ownership, may not be suppressed by improving public transit". (Kitamura, 1989)

In addition to car ownership, the gender is also included in many ridership models. Most, however, find little or only a small influence on transit ridership. For example, based on 2012's data from the American Community Survey of the U.S. Census, more women than men overall ride rail transit. People who take public transportation to work, 50.5 percent are women and 49.5 percent are male. (Goldmark, 2012)

The issue of aging U.S. population is important of all sectors of the economy, especially in public transit. National Cooperative Highway Research Program studied the effect of aging demographics of the country will impact future transit ridership. By using the data from 2001 National Household Transportation Survey, the study described that older people are less likely to travel in general and take fewer trips on public transportation than their younger counterparts. (Program, 2006)

In Cynthia's analysis of New York City's rail transit ridership, she pointed out an important dimension that socio-demographics of the visitor population have been missed in the station-level ridership studies. The New York City Region is culturally diverse and attracts a lot of visitors every year, which influencing the rail transit a lot. (Cynthia Chen, Jason Chen, James Barry, 2009)

Besides the factors mentioned above, there are many other factors influencing the rail transit ridership, such as service quality factors (Sharfuddin J. Syed, 2000), service quantity factors (John F. Kain, 1996) (Gomez-Ibanez, 1996), public finance (Sale, 1976) and so on.

## **Chapter Four: Data**

### **4.1 Where did the data come from?**

Capital Metropolitan Transportation Authority (Capital Metro) is the public agency responsible for providing mass transit service within the City of Austin and the surrounding communities. Capital Metro is interested in understanding the demographic profile and travel patterns on its various bus routes and for its overall system in order to assist them in planning their routes for the future. This study's data was based on Capital Metro 2010 on-board survey. The survey was processed to locate origins and destinations of transit trips in the new traffic analysis zone system.

Creative Consumer Research was hired to conduct intercept interviews with riders on the Capital Metro bus and rail transit systems in order to determine riders' origins, destinations, and other pertinent demographic information. This information can help me in the endeavor of answering the correlation between spatial, demographic characteristics and the rail transit ridership.

Additionally, a GIS dataset was collected from the Austin Government Online Sources.

### **4.2 Data description**

There are 28 questions in this survey questionnaire. Among these questions, 16 questions focus on spatial factors. The rest of questions focus on demographic characteristics. The total number of interviewees, who participated in this survey, is 405. The questionnaire is attached in the end of this thesis.

### **4.3 Data cleaning/data categorization**

The answers of questionnaire are all coded into options ranged from option 0 to option 999.

The options are defined in below tables. Table One is the spatial factors definition. Table Two is the demographic factors definition.

There are five answers that this study coded the answers into different categories.

The first question is “how many blocks did you walk to this station?” and “how many blocks will you walk from this station to your destination?” To clean the data, this study makes the boundaries in a large number of random blocks. The method of determining the boundaries is taking the block number which has a big change of ridership, while comparing with the block next to it. I cumulate the distribution of the count of answers at first. The distribution is showed in Figure 4.3.1. We can observe the big ridership changes in the answers of 2, 4 and 7 blocks. So I separate the data into five categories: “0 block” is categorized into “option 0”, “1 block” is categorized into “option 1”, “2 to 3 blocks” is categorized into “option 2”, “4 to 6 blocks” is categorized into “option 3”, “7 to 9 blocks” is categorized into “option 4” and “more than 9 blocks” is categorized into “option 5”. The new category distribution is showed in Figure 4.3.2

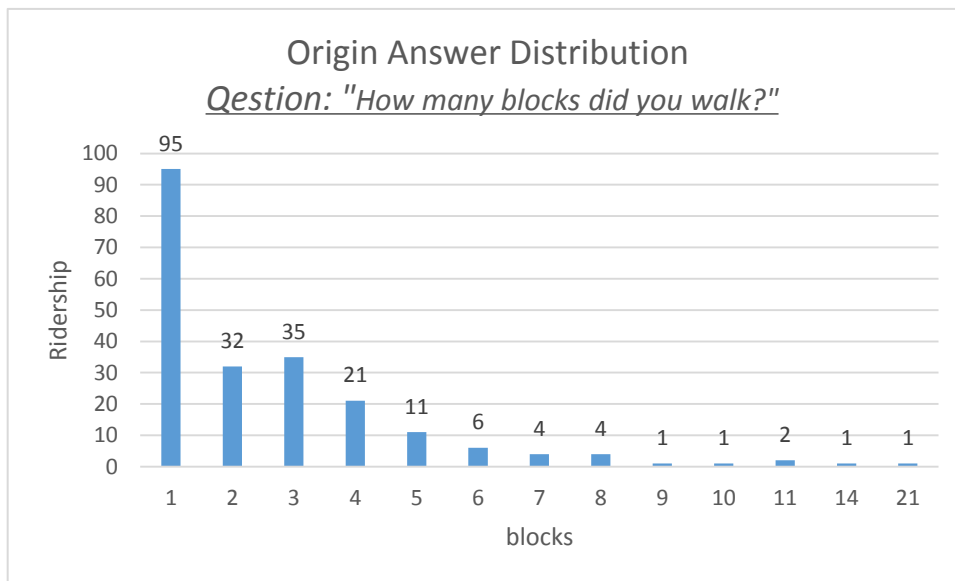


Figure 4.3.1: Walking distance (1)

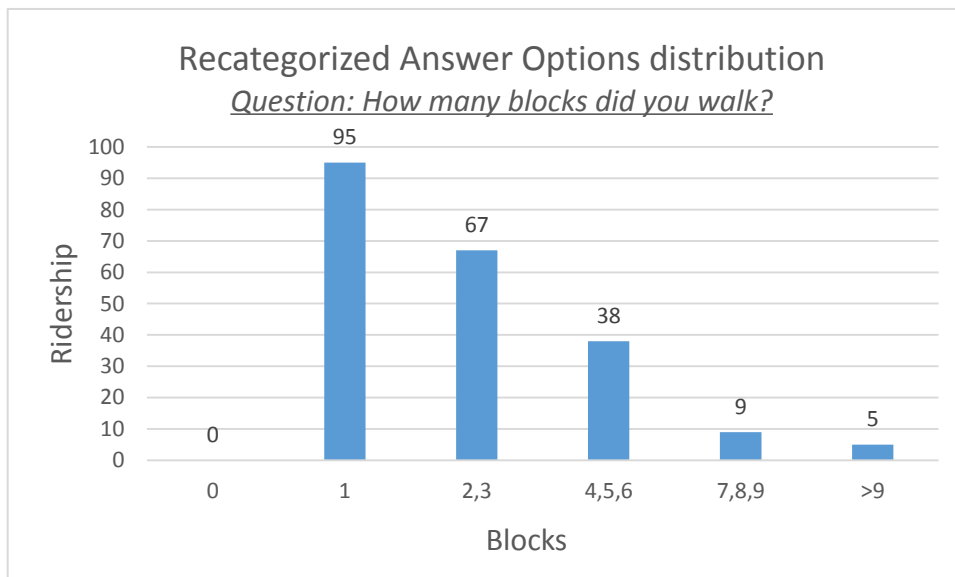


Figure 4.3.2: Walking distance (2)

The same method is also used in coding the age variable (question 16) and miles variables (question 5b, 5c). These three distribution figures and new categorized figures are showed in Figures 4.3.3, 4.3.4, 4.3.5 and 4.3.6.

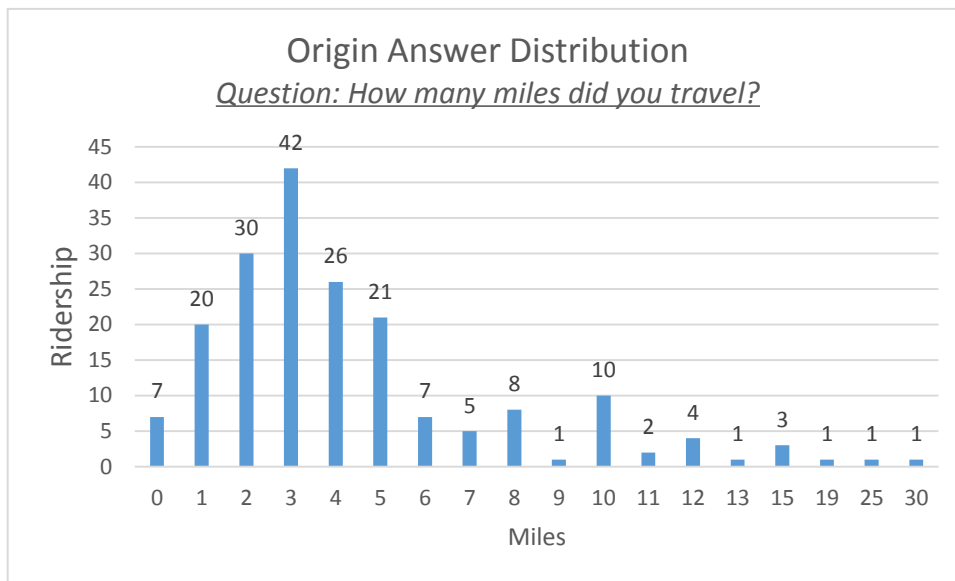


Figure 4.3.3: Traveling distance (1)

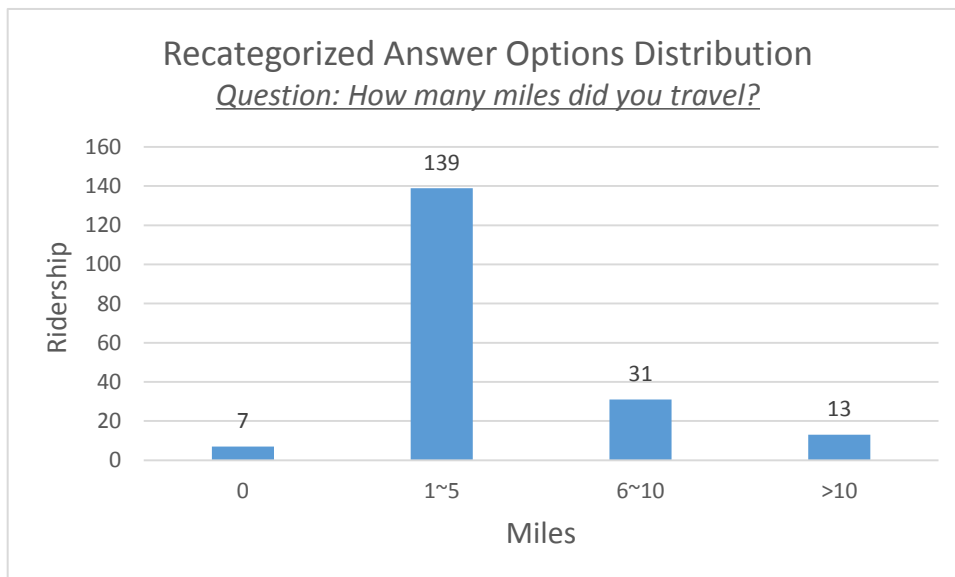


Figure 4.3.4: Traveling distance (2)

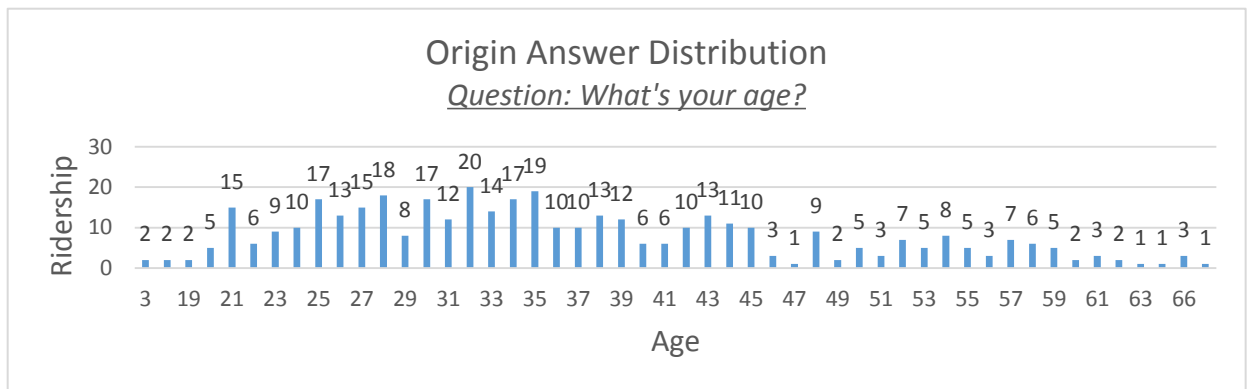


Figure 4.3.5: Age

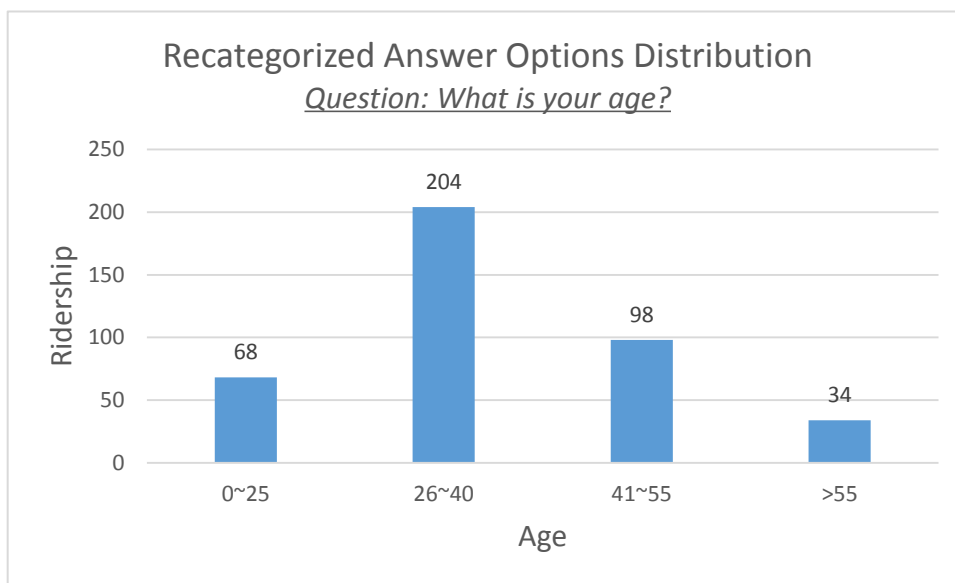


Figure 4.3.6: Categorized age option

A question for station ridership analysis is on how to define the study area. Many research defined the study area by applying the potential walking distances to a station (Kuby et al. 2004). A considerable number of studies have been conducted to evaluate walking distance to and from transit stations (Zhao and Deng 2013; Hess 2012; Guerra et al. 2011; Canepa 2007; Ewing 1999; Murray et al. 1998). In this thesis, I am not going to calculate the reasonable walking radius to and from transit station. I will let the interviewees themselves answer how long they walked from the origins to rail transit stations. Or how long are they willing to walk from the stations to destinations. I will discuss how walking distance changes will influence the



ridership. Obviously, it is reasonable that with the walking distance reduced, people are more likely to take the rail transit. This study will explore deeper in the survey data, and might bring up a good observation about the distance thresholds in ridership analysis. Especially due to the decay in walking distance, it is a possibility that ridership would be significantly changed in certain walking distance. So, above all, I keep the walking distance data intact. I'll categorize them into different options. The answers of questions are coded into different options. These option definitions can be found in the Tables 4.3.1 and 4.3.2.

QUEST	OPT 0	OPT 1	OPT 2	OPT 3	OPT 4	OPT 5	OPT 6	OPT 7	OPT 8	OPT 11	OPT 12	OPT 99
a_language		English	Spanish									
q2		Home	Work	College (Not University Texas)	The University of Texas	Shopping	Medical	Personal/Recreational	School (elementary/middle/high school)	Airport	Other	All other
q4		Transferred	Rode	Drove	Walked	Bike	Other					
q4b	one block	two to three blocks	four to five blocks	six to eight blocks	more than eight blocks							
q4c	zero mile	one to five miles	six to ten miles	more than ten miles								
q5		Transfer	Ride	Drive	Walk	Bike	Other					
q5b	zero block	one block	two to three blocks	four to six blocks	seven to nine blocks	more than nine blocks						
q5c	zero mile	one to five miles	six to ten miles	more than ten miles								
q6		Home	Work	College (Not University Texas)	The University of Texas	Shopping	Medical	Personal/Recreational	School (elementary/middle/high school)	Airport	Other	All other

Table 4.3.1: Spatial options definition

QU EST	OP T 0	OPT 1	OPT 2	OPT 3	OPT 4	OPT 5	OPT 6	OP T 7	OP T 8	OP T 9	OPT 10	OP T 11	OP T 12	OP T 13	OP T 14	OPT 15	OP T 99
q8		Cash	31 Day Metr o Pass	31 Day Expre ss Pass	Day Pass	UT Student/F aculty Pass	Mobi lity Impa ired Pass	City of Aust in	Free	Oth er	One- Zone ticket	Met ro Plus Day Pass	All zone s tick et	31- Day Met ro Plus Pass	5-D ay Met ro Plus Pass	ACC Student/F aculty Pass	All othe r
q9		Adult	Senio r	Stude nt	Child	Disabled	Don't know										
q10	Zero	One	Two	Three	Four	Five	Six										Refu sed
q11		Yes	No	Don't know													
q12		One	Two	Three	Four	Five	Six	Sev en	Eigh t	Nin e							
q13		White/ Anglo	Afric an Amer ican	Hispa nic	Asian	Native America n	Othe r										
q14		English	Spani sh	Mand arin Chine se	Vietna mese	Other	Refu sed										

Table 4.3.2: Demographic characteristic options definition

QUEST	OPT 0	OPT 1	OPT 2	OPT 3	OPT 4	OPT 5	OPT 6	OPT 7	OPT 8	OPT 9	OPT 10	OPT 11	OPT 12	OPT 13	OPT 14	OPT 15	OPT 99
q15	0-25	26-39	40-55	55+													
q16		Male	Female														
q17		\$0 - \$4,999	\$5,000 - \$9,999	\$10,000 - \$14,999	\$15,000 - \$19,999	\$20,000 - \$24,999	\$25,000 - \$29,999	\$30,000 - \$39,999	\$40,000 - \$59,999	\$60,000 - \$69,999	\$70,000 - \$79,999	\$80,000 - \$100,000	Over \$100,000	Refused (below \$20,000)	Refused (above \$20,000)	Refused	
q18		6-7 days a week	5 days a week	3-4 days a week	1-2 days a week	1-2 days a month	Less than 1 day a month	This is my first time	Don't know/refused								
q19		Less than one year	1-2 years	2-3 years	3-4 years	4-5 years	5-6 years	6-7 years	7 or more years	Refused							

Table 4.3.2: Demographic characteristic options definition

## **Chapter Five: Modeling and Analysis**

### **5.1 Method of modeling**

My study of rail transit ridership was, based on sets of quantitative and qualitative data from interviews with riders, to identify factors believed affecting ridership. Based on previous literature, studies of transit ridership have examined a wide range of factors thought to influence ridership. I group the factors from the Capital Metro's data into two categories: Socio-economic factors, and spatial factors.

#### **Social-economic factors**

For the social-economic factors, this study identified six main variables: race, age, income, gender, car-ownership, years lived in Austin. Each variable represents a question in the survey, which contains several options. Then, the study categorized these options into sub-variables. For example, based on the question of races, there are six options: White/Anglo, African, American Hispanic, Asian, Native American and Other. The categorical options will be counted as an independent variable in the regression model. So, there are six independent variables about race in this study's regression models. And the value of every variable is the percentage of ridership of the station. For example, there are 100 riders starting their trips from the Downtown station. Among them, 50 riders' race is white. So the value of Downtown station's "white race" variable is  $50/100=0.5$ . A similar method is also applied to other social-economic options.

#### **Spatial factors:**

There is a large and growing body of research examines the relationships between transportation systems, land use and urban form, and travel behavior. Policy makers and planners have some direct control over land use and the deployment of transportation systems, but less control over many of the socio-economic factors discussed above. Spatial factors are defined as the environment characteristics in certain miles from the station. (Robert Cervero, Kara Kockleman, 1997)

Actually, in some situation, if we collected certain spatial factors just in certain miles from stations, we might lose some important findings. The distant decay might cause significant changes of transit ridership, especially in a certain mile changes (Javier

Gutiérrez, Osvaldo Daniel Cardozo, Juan Carlos García-Palomares, 2011).

In fact, many travel behaviors have close correlation effect between environment and socioeconomics. This study tries to explore certain spatial factors, which I assumed are causally related to rider's travel habit and spatial environment. So, this study selected these spatial factors based on the survey questions which were focused on travel distance. For example: what origin and destination of riders' trip are; how the riders arrived the rail stations and how they will travel from this train station to destinations; the distance between origins and stations/stations and destinations; their preferences of walking distances and bus routes.; Because of the reason I just mentioned that I didn't assume a predetermined number to be my study radius. So, in this study I will not delete the samples collected from those who travelled from far away to take the rail, or those who would travel a long distance from the rail station to his/her destination.

The answer of questions, such as "how long did you travel to this rail station" and "how long will you travel from the rail station to your destination," were opened to interviewees.

There are ways to calculate the distance between people's origins and destinations to rail stations. In fact, distance measurement methods can influence the ridership forecast and analysis. (Javier Gutierrez, Juan Carlos Garcia-Palomares, 2008) In the model of this study, instead of calculating the Euclidian distance or network distance manually, I use the distance data which is offered by interviewees directly. Because my research question focused on people's travel behavior/habit, I would like to let riders to define how much distance they traveled to the rail station or how long they are willing to travel after getting off the train.

### **Valuables Correlation**

Even though I separate every survey question's options into several independent variables, we still have the persistent problem of multicollinearity. A high degree of correlation between independent variables in the same model is always a big problem in the previous research. In this study, I focused on the correlation happens between the variables coming from different questions. There can also be serious endogeneity problems between survey supply variables and transit demand. To analyze how those variables could influence each other, correlation tables were created for the two final models we obtained in the end.

## 5.2 Multiple Linear regression model

This study uses a multiple linear regression attempting to model the relationship between explanatory variables and transit rail ridership, by fitting a linear equation to observed data. Every value of the independent variable ( $x$ ) is associated with a value of the dependent variable ( $y$ ).

Formally, the model for multi-linear regression, give  $n$  variables is:

$$y = \sum_i^n x_i$$

$y$ : Station level ridership;

$x_i$ : Variables affect ridership;

$i$ : The order of variable;

$n$ : Number of variables.

While modeling, I use R-square and Akaike information criterion (AIC) to select the models, which are used in the analysis chapter.

**R-square:** In general, a model fits the data well if the differences between the observed values and the model's predicted values are small and unbiased. R-squared is a statistical measure of how close the data are to the fitted regression line. In this study it is known as the coefficient of multiple determination for multiple regression.

R-squared = Explained variation / Total variation.

R-squared is always between 0 and 1:

0 indicates that the model explains none of the variability of the response data around its mean.

1 indicates that the model explains all the variability of the response data around its mean.

Generally, the larger R-square represents the better fit of the regression model.

### **Akaike information criterion (AIC)**

This study's sample size is very small. There are only one route and nine rail stations in Austin now. And my research question focused on the station level ridership analysis. So, my sample size is nine. Obviously this data size is small. If we use ordinary regression method, for example applying R data analysis software, we can't get the reasonable output. So, I apply ridge regression method in the modeling.

The main process of modeling in this study is separated into two steps. The first step is to remove all features that are useless or highly related to some other features. I applied the M5' method, which uses Akaike information criterion (AIC) as a measurement of model quality to perform feature selection.

For any statistical model, the AIC value is:

$$AIC = 2k - 2\ln(L)$$

Where  $k$  is the number of parameters in the model, and  $L$  is the maximized value of the likelihood function for the model.

In the modeling, the first step is by giving a set of candidate models with different variables, the preferred model is the one with the minimum AIC value. Hence, AIC not only rewards goodness of fit, but also includes a penalty that is an increasing function of the number of estimated parameters. The penalty discourages over-fitting, since increasing the number of parameters in the model almost always improves the goodness of the fit.

However, the variables kept in the final models are not the factors that are closely related to the ridership. And, the variables that not involved into the models don't mean the variables is not important in changing the rail transit ridership. The models that I finally produced is built by the balance of AIC index. The lower AIC may represent the high score of complexity and high score in goodness of fit. The R-square can tell us whether the model has a good fit or not. There is no strict line of the modeling for R-square telling us which model is acceptable. In social science disciplines, R-square always not very high, especially when the sample size is not enough. Usually we can accept a model whose R-square is higher than 0.5.

So, the variables kept in final models are the factors that are closely related to the ridership. And, the variables that not involved in the models don't mean the variables are not important in changing the rail transit ridership. The final two models listed in this thesis could be a best fit of rail ridership, based on the survey data. There are some limitations of these models, I will discuss it in the following Chapters.

## **Demographic model**

There are a lot of studies, no matter based on quantitative or qualitative methods, to prove the social-economic characteristics have significant influence on ridership.

I used WEKA statistic software to build a multiple linear regression model. After



inputting the whole dataset of variables, many variables showed no influence on ridership. So, I deleted the variables who had no significant influences. The final demographic socioeconomic model output is;

$$y = \sum_i^n \alpha_i x_i$$

The final demographic model is presented in Table 5.2.1

<b>Demographic model</b>		
<b>Coefficient(<math>\alpha_i</math>)</b>	<b>Variables (<math>x_i</math>)</b>	<b>Meaning</b>
-20.8061	10.1	one car (car ownership)
15.9207	10.2	two car (car ownership)
133.2764	10.4	four cars (car ownership)
1654.9159	10.5	five cars (car ownership)
1654.916	10.6	six (car ownership)
1654.916	10.9	nine (car ownership)
-13.7102	13.3	Hispanic (race)
89.9688	13.4	Asian (race)
167.6212	13.6	Other (race)
54.3611	15.3	55+ (age)
-35.3531	17.4	\$15,000 - \$19,999
-43.1031	17.6	\$25,000 - \$29,999
102.0831	17.10	\$70,000 - \$79,999
33.6403	17.11	\$80,000 - \$100,000
50.0154	17.12	Over \$100,000
143.8583	17.15	Refused
118.0727	19.1	Less than one year (years in Austin)
-90.5405	19.3	2-3 years
-29.3324	19.4	3-4 years
35.4411		<b>Constant</b>

Table 5.2.1: Demographic model output

Correlation coefficient	0.8564
Mean absolute error	8.4136
Root mean squared error	17.2477
Relative absolute error	31.3226%
Root relative squared error	52.2899%
R-square	0.7334

Table 5.2.1: Demographic model output

## **Social-economic factors:**

### **Car ownership**

In the demographic final model we can observe that there are six variables out of 15 variables focusing on car ownership.

By comparing different coefficients of car ownership variables, the fewer the car number, the less the ridership will be. This is an obvious conclusion that we can observe from the model.

Another observation of the model is that the one car ownership variable's coefficient is negative. We could make an assumption that the household that owns one car would not likely take the rail for transit. The probable reason is the income of household with one car ownership might be relatively low. And, rail transit is more expensive than bus transit. This group might prefer to ride buses for transit. The previous research documented racial and ethnic differences in car ownership rates along with how car access differs by household income relative to the poverty line. (Alan Berube, 2006) (Cirillo, 2010). The literature also suggests that there are sizable population don't have access to household car ownership. For the less car ownership household, it is a possibility that they would choose to take bus instead of rail for transit. It is because the bus transit is cheaper than rail transit. The article that I referred to is based on the nation-wide data. However, every city's demographic, socioeconomic characteristics and travel behavior are different. It is hasty to directly apply previous research's correlation conclusion to the Austin study. So, I will test these assumptions in demographic model by using its variable correlation table. This part will be discussed in the Correlation section.

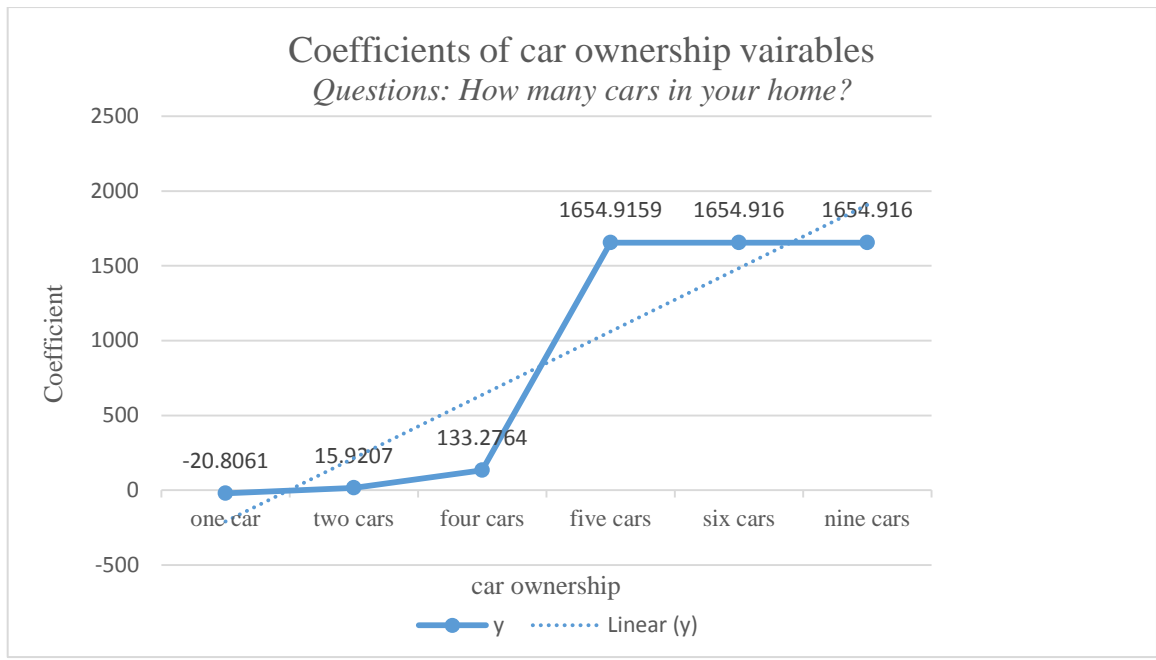


Figure 5.2.1: Car ownership

### Income

Except the option of “refused,” income does have a significant influence on ridership. Based on different ranges of income, the model shows the income ranged from \$15,000 to \$19,999 and \$25,000 to \$29,999 has negative coefficients, which are -35.3531 and -43.1031, respectively. On the contrary, the income ranged from \$70,000 - \$79,999 has the highest positive coefficient, which is 102.083.

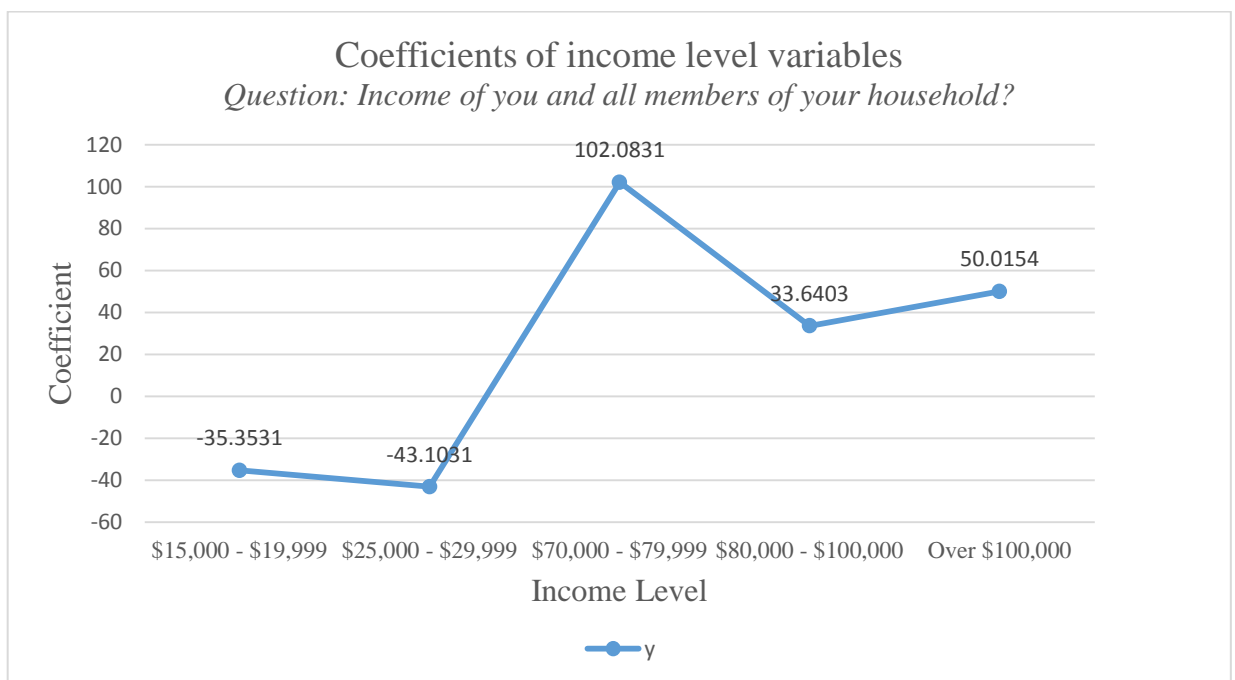


Figure 5.2.2: Income

We can make an assumption that ridership will be increased, if the income grows. High income groups are more willing to transit by rail, comparing with the low income groups. Indeed, much research has already proved, the income has a positive influence to rail transit ridership. (Spillar Robert J., G. Scott Rutherford, 1998) There are also some cases argue that the income does not act as a significant factor in influencing the rail transit ridership. For example, Gomez-Ibanez(1996) finds that ridership in Boston between 1970 and 1990, regional employment is more significant than income. (Gomez-Ibanez, 1996)

### **Race**

In the race variable category, after modeling selection, there are two race options left. One is “Asian,” The other one is “Other races.” Asian variable coefficient is 108.9199. Asian probably would like riding rail transit. The reasons of this observation could be various. More research and data are required.

### **Years have been stayed in Austin:**

In this category, there are two variables selected in the final spatial model as relatively significant influence variables. The valuable, which has largest absolute coefficient: 118.0727 in this category, is people lived “less than one year” in Austin. And, the variable of “two to three years” has a negative coefficient: -90.5405. “Three to four years” variable’s coefficient is -29.3324. Based on the model output, it is reasonable to make a conclusion that people who live less than one year in Austin are more willing to take rail. It is possible that a large number of visitors randomly participated the survey, so the data collected from them could increase the positive coefficient of the “live less than one year” valuable. Because this survey didn’t ask the interviewees information of whether they are visitors or native residents, this study won’t know the percentage of people who are visitors. In factor, visitors will affect a lot in transportation, especially in transit ridership. For example Cynthia Chen (2009) analyzed the diurnal pattern of subway ridership. The study pointed out an important dimension that socio-demographics of the visitor population have been missed in the station-level ridership studies. (Cynthia Chen, Jason Chen, James Barry, 2009) The New York City Region is culturally diverse, and attracts a lot of visitors every year. However, things are different in Austin. We can’t get the visitor’s socioeconomics through this survey data, but we can get a conclusion that more people with short period living experience in Austin may be more willing to take transit by rail.

## **Gender**

In this model, it doesn't show gender influencing the rail transit ridership. But, based on the literature, gender does have a difference. For example, Los Angeles Metro's 2012 Customer Satisfaction Survey focuses on the issue of ridership by gender. It proved more women than men overall ride Metro although there are notable exceptions: higher-income riders tend to be male, as are riders with a car available to them. (Sullivan, 2013).

But, why there are no gender variables in the model? There are probably three reasons. One reason is that my model is not critical enough. Because the limitation of sample size, under-fitting might be an issue in my model. The second reason is Austin's situation can't approve the theory which other studies observed. Last but not least, it might be because of the modeling method. I use the m5 method, and AIC will delete one or two variables, which has a very close correlation with each other. To test the third one, I create the full size of demographic correlation table later. The correlation coefficient is very high between income and gender. We still can conclude that gender actually has a significant influence on ridership.

## **Spatial model**

Table 5.2.2 is the final model of spatial variables.

<b>Spatial Model</b>		
Coefficient	Variables (the numbers are right question number)	Meaning
2096.544	2.8	School (elementary/middle/high school)
-40.3975	4.5	Bike
77.0543	4b.5	More than 9 blocks
26.0505	4c.2	6~10 miles
101.6137	5b.0	0 blocks
-8.4148	5b.1	1 blocks
18.5795	5b.2	2~3 blocks
-78.7762	5b.4	7~9 blocks
71.1441	5c.0	0 miles
-36.5012	5c.1	1~5 miles
18.5737	5c.2	6~10 miles
60.3501	5c.3	> 10 miles
652.3222	6.3	College (not University Texas)
174.3706	6.5	Shopping
1278.0015	6.8	School (elementary/middle/high school)
54.4832		Constant

Table 5.2.2: Spatial model output

Correlation coefficient	0.927
Mean absolute error	9.306
Root mean squared error	11.745
Relative absolute error	34.645%
Root relative squared error	35.608%
R-square	0.859

Table 5.2.2: Spatial model output

### **Spatial:**

Based on the AIC and R-square, there are mainly three kinds of variables kept in the final spatial model. The Origin/Destination types, travel mode between Origin/Destination and stations, distance between Origin/Destination and stations;

#### **Origin / Destination**

In the survey, Capital Metro defined six options for “where did you from?” and “where will you go?” questions. They are “Home”, “Work”, “College (Not University Texas)”, “The University of Texas”, “ Shopping”, “Medical”, “Personal/Recreational”, “School (elementary/middle/high school)”. The coefficients of Origin and Destination options are bigger than other questions’ options.

When answering “where will you go?” a large percent of riders’ destination will give the answer “School (elementary/middle/high school).” It is just as same as the origin trend. And, the coefficient of this destination option is also quite big: 1278.0015. Additionally, the destination of “College (not University Texas)” comes up with third biggest positive coefficients among the whole spatial variables, which is 652.3222. The last variable of trip purpose in the final model is “Shopping;” its coefficient is 174.3706, and also ranked the fourth biggest coefficient.

We could make a conclusion that trip purpose significantly affects the rail transit mode choice in Austin. Especially, when the number of school trips increases, rail transit ridership will increase significantly.

Why “school” variable coefficient is far greater than others? On one hand, no matter in the origin or destination analysis, there is no literature gave a strong evidence to approve people aged 0 to 18 is such important in the rail transit ridership. On the other hand, in my demographic model, only “55+ aged” option left as the independent variable, which is selected from the age socioeconomic variables. Some literature mentioned that the combination of trip purpose and walking distance can influence the transit ridership. For example, a study conducted in Atlanta proved self-identified rail riders primarily arrive rail station to within convenient walking distance of employment, such as in the central business district (CBD). (Jeffrey Brown, Gregory Thompson, Torscha Bhattacharya, Michal Jaroszynski, 2013)

For these reasons, the reason could be several schools located close to the rail station. The education land use map is showed in Figure 5.2.3. Walking or travel distance is short between stations and origins/destination. So, it is convenient and safe to travel for young people in school age. This assumption can be approved in the correlation table. I will discuss it in the following Correlation Table Analysis chapter.



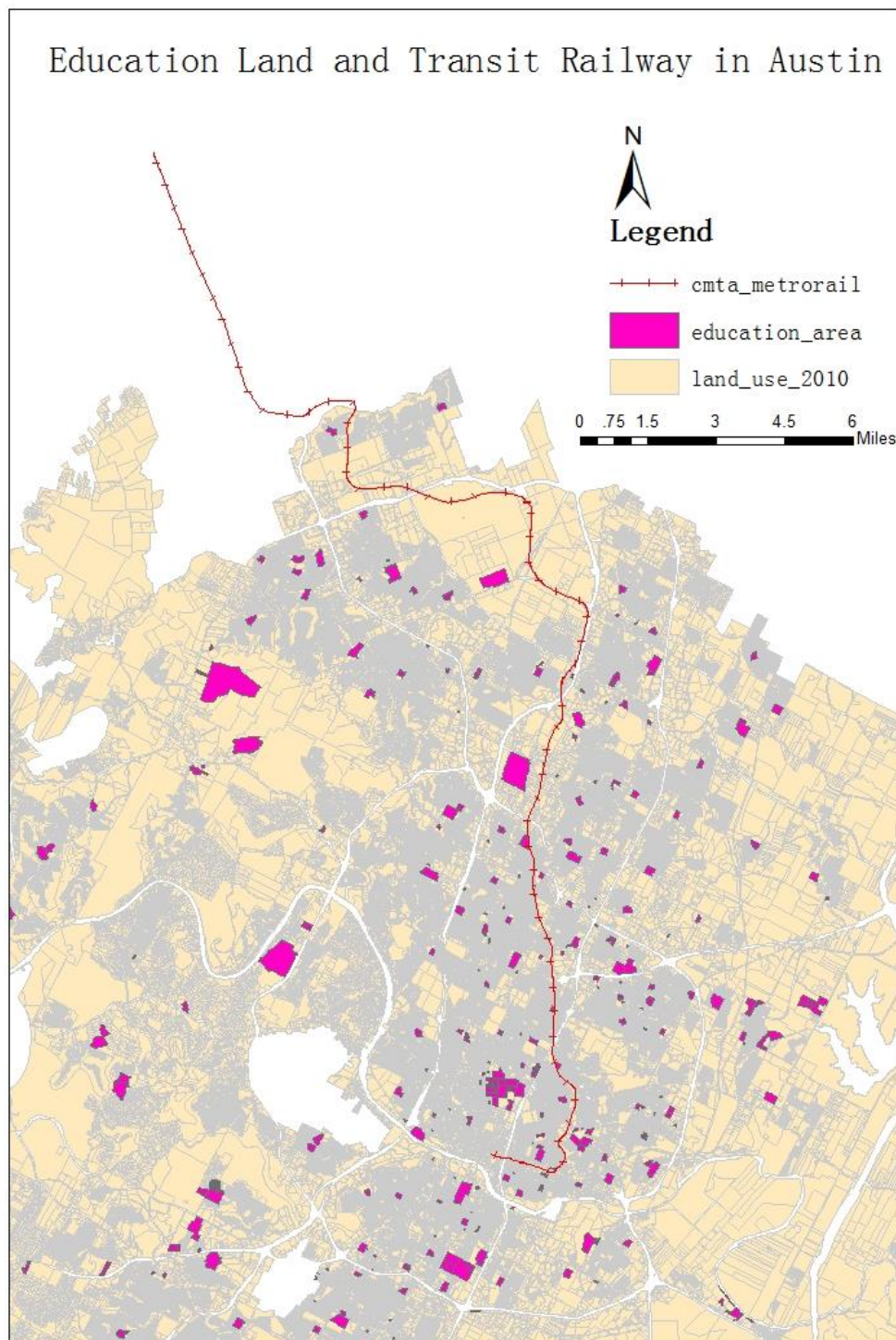


Figure 5.2.3: Education land and transit rail in Austin (Austin, 2010 )

### Travel modes choice from Origin/Destination to the rail stations

They are travel modes from Origin (Destination) to (from) rail transit stations. The travel modes between Origin/Destination and rail transit stations are categorized into

five modes: “Transferred”, “Rode”, “Drove”, “Walked”, “Bike” and “Other.”

Through the modeling, we can make an assumption that biking people do not like rail. Because the coefficient of “bike” mode is negative, other travel modes may not have significant influences on rail transit ridership, on the way from origins to stations. There is no primary variable of the travel mode from stations to destinations.

### Walking distance

Capital Metro designed two questions about the walking distance of riders. One question is “how many blocks did you walk from origins to stations?” The other one asks how many blocks from stations to destinations, by feet. The data shows a lot of people were willing to walk much more blocks from origins, but not willing to walk very long from stations to destinations.

It is an interesting observation. We could probably make an assumption that people are more willing to walk longer at the beginning of the trips. Because for the range of 7~9 blocks walking distance, variable coefficient from origin to destination is positive: 77.0543. But, variable coefficient from station to destination is negative: -78.7762. When we focus on the answer of origin to station, there are two options having negative coefficients. They are “1 block” and “7~9 blocks”. In the Figure 5.2.4 it is easy to observe the trend of coefficient, which is reduced while blocks increasing. Additionally, “one block” comes up with a negative coefficient. It gives us a possibility that if people need to walk one block from rail transit station to destination, they prefer to use other travel mode.

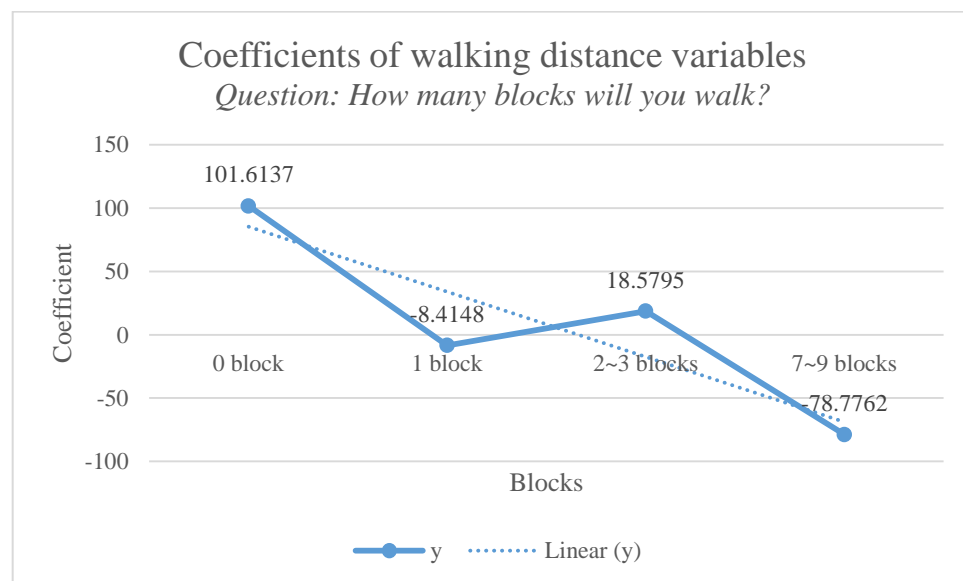


Figure 5.2.4: Walking distance (3)

### **Travel distance**

Besides the walking distance, there are also two questions focus on the total distance that people are willing to travel, between Origin/Destination and stations. Based on the model output, the ridership would not be significantly influenced by the distance between origin and station. On the contrary, travel distance between destination and station could influence the ridership significantly. Below is Figure 5.2.5. It talks about the various mile ranges' coefficients.

Same as the walking distance analysis, we can use it to analyze the travel distance from stations to destinations. Obviously, different from the walking distance, the model doesn't give out a smooth and clear line of ridership trend. If the destination is in a short distance from stations, such as 0 or one mile, people are willing to take the rail for transit. However, if the distance ranges from 1 mile to 5 miles, it would probably cause the decrease of rail transit ridership. But, long travel distance, for instance more than 6 miles, will actually increase the ridership again. This observation of preference on travel distance might have a high correlation with the travel destination type and travel mode. For example, if travel distance from station to destination is too long, people will prefer to take auto or transfer to arrive final destination. The travel distance between 1 mile and 5 miles, is an embracing distance. It is Because it is too short for bus and auto mode, but too long for walking. Riding bicycle sometimes is not a convenient option for public. So, people, whose travel distance dropped in the "embracing travel distance", might even choose other transit mode instead of rail way. To test my assumption, I analyze the correlation between walking distance and travel mode in the following chapter..

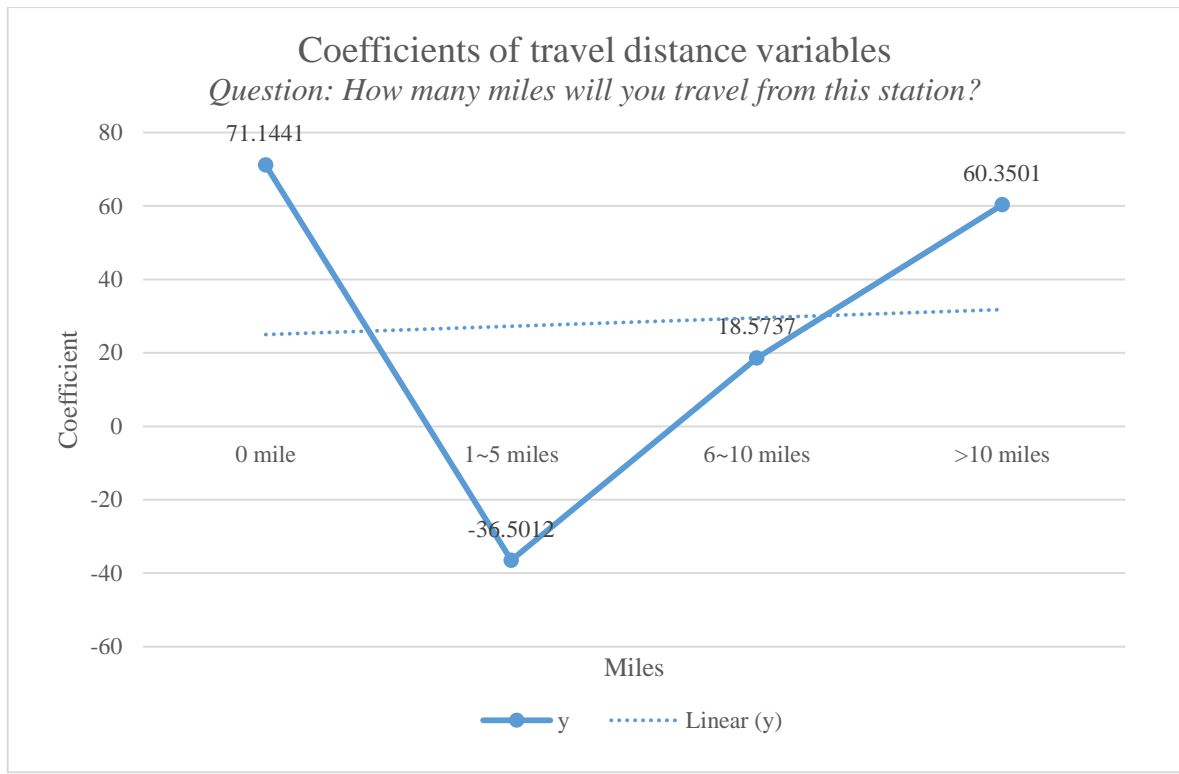


Figure 5.2.5: Travel distance (4)

## 5.3 Correlation Table Analysis

Correlation variables are always a problem for ridership regression analysis. So I create a correlation table for each final model. By closely examining correlation coefficients in these two table, we can approve or reject the assumptions that I gave in the previous analysis.

Two models' correlation tables can be found in Table 5.3.1 and Table 5.3.2.

The variables are all named by the pattern of “question code.option code”. For example the first question's first option is named “1.1”. The definition of variables can be find in the Table 4.3.1 and Table 4.3.2.

The first assumption is that one car ownership households are closely correlated to the low household income. Overall, two variables in the final demographic model can be counted as low income factor. One is 17.4 (\$15,000 - \$19,999), the other one is valuable 17.6 (\$25,000 - \$29,999). The one car ownership variable has a high positive correlation value with the valuable 17.6. The number is 0.59. Variable 17.4 has almost no influence to one car ownership, since the correlative value is -0.01. High income

variables are all showing negative effect on the one car ownership. Based on the demographic model, lower income brings less rail transit ridership. And lower income brings more household drop into one car ownership group. All in all, it is reasonable to say that one car ownership's increment could also reduce the rail transit ridership. The conclusion of car ownership in the demographic model analysis section is confirmed again.

Second assumption is about the travel distance. Looking at the spatial final model, it is easy to observe people are willing to walk/travel longer from origin to station, but not from station to destination. In the correlation table, there is no evidence that shows which group of riders willing to travel longer distance before arriving at the station.

	10.1	10.2	10.4	10.5	10.6	10.9	13.3	13.4	13.6	15.3	17.4	17.6	17.1	17.11	17.12	17.15	19.1	19.3
10.1	1																	
10.2	-0.15	1																
10.4	0.59	0.44	1															
10.5	-0.38	0.12	0.21	1														
10.6	-0.38	0.12	0.21	1	1													
10.9	-0.38	0.12	0.21	1	1	1												
13.3	-0.04	-0.61	-0.51	0.07	0.07	0.07	1											
13.4	0.14	0.32	0.2	0.21	0.21	0.21	-0.01	1										
13.6	-0.56	0.8	-0.08	0.1	0.1	0.1	-0.5	0.29	1									
15.3	0.02	0.22	0.56	0.24	0.24	0.24	-0.72	-0.21	0.16	1								
17.4	-0.01	-0.61	-0.42	-0.05	-0.05	-0.05	0.78	-0.49	-0.55	-0.45	1							
17.6	0.59	0.25	0.49	-0.07	-0.07	-0.07	-0.1	-0.04	-0.05	0.04	0.16	1						
17.1	-0.46	-0.04	-0.32	0.15	0.15	0.15	-0.05	0.52	0.35	0.1	-0.41	-0.74	1					
17.11	-0.36	0.11	-0.29	0.14	0.14	0.14	0.05	0.46	0.54	0.11	-0.22	-0.17	0.7	1				
17.12	-0.44	0.89	0.12	0.12	0.12	0.12	-0.65	0.21	0.9	0.19	-0.68	-0.01	0.14	0.18	1			
17.15	-0.8	0.54	-0.03	0.48	0.48	0.48	-0.38	0.01	0.64	0.3	-0.36	-0.5	0.39	0.14	0.68	1		
19.1	-0.5	0.36	0.09	0.34	0.34	0.34	-0.33	0.15	0.58	0.6	-0.42	-0.37	0.63	0.69	0.4	0.63	1	
19.3	0.65	0.01	0.21	-0.44	-0.44	-0.44	0.2	0.35	-0.11	-0.23	0.19	0.61	-0.18	0.24	-0.3	-0.68	-0.21	1
19.4	-0.57	-0.23	-0.7	-0.16	-0.16	-0.16	0.25	-0.34	-0.01	-0.48	0.37	-0.54	0.08	-0.3	0.02	0.33	-0.28	-0.48

Table 5.3.1: Correlation table of demographic model

	<b>2.8</b>	<b>4.5</b>	<b>4b.5</b>	<b>4c.2</b>	<b>5b.0</b>	<b>5b.1</b>	<b>5b.2</b>	<b>5b.4</b>	<b>5c.0</b>	<b>5c.1</b>	<b>5c.2</b>	<b>5c.3</b>	<b>6.3</b>	<b>6.5</b>	<b>6.8</b>
<b>2.8</b>	1														
<b>4.5</b>	0.03	1													
<b>4b.5</b>	0.26	-0.13	1												
<b>4c.2</b>	0.07	-0.58	0.58	1											
<b>5b.0</b>	0.06	-0.49	0.33	0.74	1										
<b>5b.1</b>	-0.14	0.43	-0.17	-0.66	-0.72	1									
<b>5b.2</b>	0.27	-0.48	-0.05	0.35	0.51	-0.87	1								
<b>5b.4</b>	-0.24	0.67	0.13	0.09	-0.04	0	-0.33	1							
<b>5c.0</b>	0.49	0.36	-0.05	-0.31	-0.28	0.23	-0.18	0.21	1						
<b>5c.1</b>	-0.59	-0.21	0.06	-0.12	-0.44	0.3	-0.21	-0.09	-0.2	1					
<b>5c.2</b>	0.31	-0.08	-0.17	0.34	0.48	-0.56	0.39	0.02	-0.09	-0.81	1				
<b>5c.3</b>	0.33	0.32	0.2	-0.14	0.25	0.2	-0.12	0	-0.08	-0.48	-0.01	1			
<b>6.3</b>	0.24	-0.64	0.62	0.8	0.61	-0.45	0.36	-0.13	-0.07	0.09	-0.11	0.06	1		
<b>6.5</b>	0.88	0.23	0.38	0.02	0.06	-0.04	0.08	0.07	0.75	-0.5	0.1	0.34	0.26	1	
<b>6.8</b>	0.51	-0.23	-0.04	0.27	0.55	-0.43	0.57	-0.14	0.14	-0.49	0.24	0.48	0.52	0.44	1

Table 5.3.2: Correlation table of spatial model

## Chapter Six: Conclusion

In this thesis I have present two ridership analysis model for Austin. One is focusing on the socioeconomic characteristics, such as car ownership, race, income, age, years living in Austin. The other one is focusing on the spatial factors, such as trip purpose, trip origin, walking distance and additional travel distance to/from stations. The models can be recommended applying to analysis the potential factors, which could influence the rail transit ridership.

Based on R-square, demographic model has a better fit than spatial model. It is not a surprise that R-square is larger in demographic model than in spatial model. To summarize the results obtained we can say that, for Austin, socioeconomic factors influence the rail transit ridership more than spatial factors.

Low income incensement produces significant decrease in the railway ridership. High income groups are much more likely to ride rail transit than low income groups.

Car ownership is also an important variable in the ridership analysis. More cars owned in a household, will likely to take more rail trips for transit. However, income has a strong effect on car ownership, while other socioeconomic factors of riders produces relatively small effects.

One of the important observations from Austin is: gender doesn't influence rail transit ridership. Ordinary, age doesn't influence a lot on rail transit ridership. However, compared to other age groups, people in Austin older than 56 years are tending to take the more rail trips for transit.

School purpose trip has a significant influence on ridership. The same effect exists in the trip starting from school. People are willing to walk longer from origin to stations than from station to destination. And travel distance doesn't act as an important role on ridership before people arrive the station. But different travel distance, after people leave the station, will bring different effect on ridership.

The model can not only provide alone analysis, but also can be integrated into the citywide model or into Capital Metro rail transit analysis models. This can enhance the complexity of modeling systems since this modeling method can be used in the small sample size case. This modeling method testing new rail transit scenarios influences route making decisions.



## **Chapter Seven: Limitation and future work**

### **7.1 Data limitation:**

The biggest limitation of this study's data is small size of our sample. It is why the models I developed in this study could be not performed well in the goodness of fits.

There is also a risk of bias of data collection process. Because the survey questionnaires were developed by Capital Metro for CAMPO's Austin 2035 Long Range Plan. Capital Metro has its own bias while design this survey project.

This survey is conducted in several months. The time period is too short to rule out the influence from the weather, seasons, and holiday. The dataset also lack the information of participation time of interviewees. So we lack the variables of weekday and weekend, peak hour and off peak hour.

### **7.2 Model limitation:**

There is a risk that some variables in the models are not trustful. Just as mentioned in the modeling methodology, I use "m5' method" and ridge regression method. It is different from ordinary linear regression method, then I chose to use Weka data mining software in stand of R statistical data analysis software. On the one hand, Weka can't out put the p-value. My sample size is just nine, which determined the model's degree of freedom is eight. It is very hard for me to build a traditional linear regression model with less than eight variables, especially I made the assumptions that many demographic characteristics and spatial factors can influence the rail transit ridership.

### **7.3 Observation limitation**

Also the findings of this study suffer from problems of self-selection bias. The observation is limited, especially in spatial factors, such as population and employment density, traffic congestion levels, parking availability, mixed land use, sidewalk design and station design. These spatial factors are researched in many studies to explain much variation in transit ridership.

Additionally, among the variables I finally keep in my models, there could be the colinearity among these variables. Even though the modeling method in my study could select one variable and get rid of the others, which have high colinearity with the selected one. The spatial variables colinearity with socio-economic variables related to transit use raise questions about both the direction of cause and effect.

## **7.4 Future work**

I suggest future testing focus on more localized changes where significant effects could be observed. For example, I prefer to collect the data of population and employment density, traffic congestion levels, parking availability, mixed land use, sidewalk design and station design.

While a new rail project is constructed finished, I suggest collect the more data from the new route.

## Appendix

### Appendix A: Questionnaire

CCR #10-4188 - MetroRail  
4/14/10  
Draft # 3

#### METRORAIL INTERCEPT SURVEY

Hi, my name is \_\_\_\_\_ and I have been retained by Capital Metro to conduct a brief interview regarding your experience with riding the train. This survey will assist Capital metro to improve existing services and will only take about five minutes. We know your time is valuable so as a way of saying thank you, if you complete this survey and provide us with your name and telephone number your name will be entered into a contest to win a monthly pass. If you already ride free, you can give the pass to a family member or friend.

S1. Have you participated in an on-board survey for Capital Metro in the past 3 months?

- ( ) Yes - **Thank & Terminate**  
( ) No - Continue

a. Would you prefer to continue this survey in English or Spanish?

- ( ) English ( ) Spanish

1. What are the names of the cross streets where you got on **THIS** train? (**Please specify street, lane, road, etc... and if applicable, east, west, north, or south**)

Nearest corner: \_\_\_\_\_ & \_\_\_\_\_  
First street name Second street name

2. Where did you come from?

- |   |              |
|---|--------------|
| ( ) Home                                    | ( ) Shopping |
| ( ) Work                                    | ( ) Medical  |
| ( ) College(other than University of Texas) | ( )          |
| Personal/Recreational                       |              |
| ( ) The University of Texas                 | ( ) School   |
| (Elementary/Middle/High School)             |              |

( ) Airport  
(specify) \_\_\_\_\_

( ) Other

3. What is the address **OR** nearest corner of the place you started your journey today? (**Please specify street, lane, road, etc... and if applicable, east, west, north, or south**)

Address \_\_\_\_\_

Block Number  
Street Name

Nearest Corner \_\_\_\_\_ & \_\_\_\_\_  
First street name Second street name

IF THE RESPONDENT CANNOT PROVIDE A STREET ADDRESS, AFTER PROBING, ASK:

What is the nearest landmark or building from the place you started your journey today

\_\_\_\_\_

4. How did you get to the TRAIN STAION?

- ( ) Transferred from Bus Route #: \_\_\_\_\_. How many blocks did you walk from that bus to this one? \_\_\_\_\_ (0 or more)
- ( ) Rode with someone or be picked up – we will drive \_\_\_\_\_ miles.
- ( ) Drove my car \_\_\_\_\_ miles.
- ( ) Walked \_\_\_\_\_ blocks (0 or more).
- ( ) Rode a bike \_\_\_\_\_ miles.
- ( ) Other (specify) \_\_\_\_\_

5. How will you get from this bus to your final destination?

- ( ) I will transfer to Bus Route #: \_\_\_\_\_. How many blocks will you walk from this train to the bus? \_\_\_\_\_ (0 or more)
- ( ) I will ride with someone or be picked up – we will drive \_\_\_\_\_ miles.
- ( ) I will drive a car \_\_\_\_\_ miles.
- ( ) I will walk \_\_\_\_\_ blocks (0 or more).
- ( ) I will ride a bike \_\_\_\_\_ miles.
- ( ) Other (specify) \_\_\_\_\_

- 5b. If you will not use a bus to get to your final destination from this rail station, what is the reason?

( ) Not sure the bus will take me where I need to go

- ( ) The bus is not as convenient as my current travel method
- ( ) Weather not a factor now, will be more likely to use bus when hot or cold
- ( ) other \_\_\_\_\_

5c. Are you aware that there are connector buses at the MLK and Downtown stations?

- ( ) Yes
- ( ) No

5d. How did you plan your trip today?

- ( ) Capital Metro online trip planner
- ( ) Google Maps
- ( ) Printed Materials from Capital Metro
- ( ) Signage at rail stations
- ( ) Asked a friend
- ( ) Online materials
- ( ) Called the Go Line (Capital Metro Customer Service)

6. Where are you going to?

- |   |              |
|---|--------------|
| ( ) Home                                    | ( ) Shopping |
| ( ) Work                                    | ( ) Medical  |
| ( ) College(other than University of Texas) | ( )          |
| Personal/Recreational                       |              |
| ( ) The University of Texas                 | ( ) School   |
| (Elementary/Middle/High School)             |              |
| ( ) Airport                                 | ( ) Other    |
| (specify)_____                              |              |

7. What is the address **OR** nearest corner of your final destination? (**Please specify street, lane, road, etc... and if applicable, east, west, north, or south**)

Address\_\_\_\_\_

	Block	Number
Street Name		
Nearest Corner_____&_____		
	First street name	Second street
name		

IF THE RESPONDENT CANNOT PROVIDE A STREET ADDRESS, AFTER PROBING, ASK:

What is the nearest landmark or building to your final destination\_\_\_\_\_

8. How did you pay to get on this train?

- |   |  |
|---|--|
| <input type="checkbox"/> One-zone Ticket        | <input type="checkbox"/> All-zones Ticket        |
| <input type="checkbox"/> Mobility Impaired Pass | <input type="checkbox"/> 31 Day MetroPlus Pass   |
| <input type="checkbox"/> City of Austin         | <input type="checkbox"/> 5-Day MetroPlus Pass    |
| <input type="checkbox"/> MetroPlus Day Pass     | <input type="checkbox"/> UT Student/Faculty Pass |
| <input type="checkbox"/> Other (specify)_____   |  |

9. Which fare category do you pay

- |                                  |                                   |
|----------------------------------|-----------------------------------|
| <input type="checkbox"/> Adult   | <input type="checkbox"/> Child    |
| <input type="checkbox"/> Senior  | <input type="checkbox"/> Disabled |
| <input type="checkbox"/> Student |                                   |

10. How many working cars, trucks, or vans are available for use by your household?

- |                               |                              |                              |  |
|-------------------------------|------------------------------|------------------------------|--|
| <input type="checkbox"/> Zero | <input type="checkbox"/> One | <input type="checkbox"/> Two | <input type="checkbox"/> Three or more |
|-------------------------------|------------------------------|------------------------------|--|

11. Could you have used one of these vehicles to make **THIS TRIP** instead of riding the train?

- ☐ Yes  
☐ No

12. How many people reside in your household? (Family and non-family members)

- |                              |                              |                                |                               |                               |                              |  |
|------------------------------|------------------------------|--------------------------------|-------------------------------|-------------------------------|------------------------------|--|
| <input type="checkbox"/> One | <input type="checkbox"/> Two | <input type="checkbox"/> Three | <input type="checkbox"/> Four | <input type="checkbox"/> Five | <input type="checkbox"/> Six | <input type="checkbox"/> Seven or more |
|------------------------------|------------------------------|--------------------------------|-------------------------------|-------------------------------|------------------------------|--|

13. (RACE/ETHNICITY) Are you...?

- |                                      |   |                          |
|--------------------------------------|---|--------------------------|
| <input type="checkbox"/> White/Anglo | <input type="checkbox"/> African American | <input type="checkbox"/> |
| <input type="checkbox"/> Hispanic    | <input type="checkbox"/> Native American  | <input type="checkbox"/> |
| <input type="checkbox"/> Asian       | <input type="checkbox"/> Other_____       |                          |

(specify)

14. What is your preferred language spoken at home?

- |                                     |   |   |
|-------------------------------------|---|---|
| <input type="checkbox"/> English    | <input type="checkbox"/> Spanish              | <input type="checkbox"/> Mandarin Chinese |
| <input type="checkbox"/> Vietnamese | <input type="checkbox"/> Other (specify)_____ |   |

15. What is your age? \_\_\_\_\_ years

16. BY OBSERVATION: GENDER

- ☐ Male

(    ) Female

**(HAND RESPONDENT CARD A)**

17. Please read off the letter on this card that best represents your total combined yearly income of you and all members of your household (including non-family members living in your household).

A	\$0 - \$4,999	E	\$20,000 - \$24,999	I	\$60,000 - \$69,999
B	\$5,000 - \$9,999	F	\$25,000 - \$29,999	J	\$70,000 - \$79,999
C	\$10,000 - \$14,999	G	\$30,000 - \$39,999	K	\$80,000 - \$100,000
D	\$15,000 - \$19,999	H	\$40,000 - \$59,999	L	Over \$100,000

**IF REFUSED**    ---    Is your annual household income above or below \$20,000.

(    ) Above \$20,000    (    ) Below \$20,000    (    ) Refused

18. How often do you use Capital Metro?

(    ) 6-7 days a week	(    ) 1-2 days a month
(    ) 5 days a week	(    ) Less than 1 day a month
(    ) 3-4 days a week	(    ) This is my first time
(    ) 1-2 days a week	

19. How long have you lived in the Austin area?

(    ) Less than one year	(    ) 4-5 years
(    ) 1-2 years	(    ) 5-6 years
(    ) 2-3 years	(    ) 6-7 years
(    ) 3-4 years	(    ) 7 or more years

Thank you for your time. So that we may enter you in a contest to win a monthly or annual pass, may I have your name and telephone number?

Name \_\_\_\_\_

Telephone: \_\_\_\_\_

## Appendix B: Question Name Definitions

ID	Field	Question
1	respnum \$	*Respnum
2	intv	ENTER INTERVIEWER ID
3	route	ENTER ROUTE:
4	BLOCK	BLOCK :
5	direct	DIRECTION:
6	a_lang	A. Language:
7	s1	S1. Have you participated in a survey for Capital Metro in the past 3 months?
8	q1	1. What are the names of the cross streets where you got on THIS train?
9	Q1_STREET	Q1_STREET
10	q2	2. Where did you come from?
11	q3	3. What is the address OR nearest corner of the place you started your journey today?
12	Q3_STREET	Q3_STREET
13	q4	4. How did you get to the TRAIN STATION?
14	q4a	4a. What bus route did you transfer from?
15	q4b	4b. How many blocks did you walk?
16	q4c	4c. How many miles did you travel to this train station?
17	q5	5. How will you get from this train station to your final destination?
18	q5a	5a. What bus route will you transfer too?
19	q5b	5b. How many blocks will you walk?
20	q5c	5c. How many miles will you travel from this train station?
21	q5bb	5bb. Why will you not use a bus to get to your final destination from this rail station?
22	q5cc	5cc. Are you aware that there are connector buses at the MLK and Downtown stations?
23	q5dd	5dd. How did you plan your trip today?
24	q6	6. Where are you going to?
25	q7	7. What is the address OR nearest corner of your final destination?
26	Q7_STREET	Q7_STREET
27	q8	8. How did you pay to get on this train?
28	q9	9. Which fare category do you pay?
29	q10	10. How many working cars, trucks, or vans are available for use by your household?
30	q11	11. Could you have used one of these vehicles to make THIS TRIP instead of riding the train?



31	q12	12. How many people reside in your household? (Family and non-family members)
32	q13	13. Are you....?
33	q14	14. What is your preferred language spoken at home?
34	q15	15. What is your age?
35	q16	16. BY OBSERVATION: GENDER.
36	q17	17. Please read off the letter on this card that best represents your total combined yearly income of you and all members of your household (including non -family members living in your household).
37	q18	18. How often do you use Capital Metro?
38	q19	19. How long have you lived in the Austin area?
39	qname	Thank you for your time. So that we may enter you in a contest to win a monthly or annual pass, may I have your name?
40	qphone	And telephone number?
41	start	START
42	ddate	DDATE

# Bibliography

- Alan Berube, E. D. (2006). Socioeconomic Differences in Household Automobile Ownership Rates: Implications for Evacuation Policy. *Earlier Faculty Research-University of California Transportation Center*.
- Alan T. Murraya, Rex Davisb, Robert J. Stimsonc, Luis Ferreirad. (1998). Public Transportation Access. *Transportation Research Part D: Transport and Environment* 3(5), 319-328.
- Austin, C. o. (2010 ). *GIS map download*. Retrieved from AustinTexasGov, the official website of the City of Austin: [ftp://ftp.ci.austin.tx.us/GIS-Data/Regional/coa\\_gis.html](ftp://ftp.ci.austin.tx.us/GIS-Data/Regional/coa_gis.html)
- Brain D. Taylor, Camille N.Y. Fink. (2001). *The Factors Influencing Transit Ridership: A Review and Analysis of the Ridership Literature*. Los Angeles: UCLA Institute of Transportation Study.
- Brinckerhoff, P. (1996). *Transit and Urban Form, TCRP Report 16, vol. 1*. Washington, DC.: Transportation Research Board, National Research Council.
- Cervero, R. (2004). *Transit Oriented Development in America: Contemporary Practices, Impacts, and Policy Directions. International Planning Symposium on Incentives, Regulations, and Plans – the Role of States and Nation-States in Smart Growth Planning*. University of Maryland.
- Cervero, Robert. (2006). Alternative approaches to modeling the travel-demand impacts of smart growth. *American Planning Association*, 285.
- Chu, X. (2004). *Ridership Models at the Stop Level*. University of South Florida: National Center for Transit Research .
- Cirillo, C. (2010). *Automobile Ownership Model*. Washington D.C.: The National Center for Smart Growth Research and Education at the University of Maryland.
- Connect, P. (2014). *Project Connect Proposed Map*. Retrieved from Project Connect: <http://www.projectconnect.com/sites/default/files/Project%20Connect%20System%20Plan%20FINAL.pdf>

- Coorporation, C. A. (2013). *Austin Urban Rail*. Retrieved from Central Austin CDC:  
[http://centralaustincdc.org/transportation/austin\\_urban\\_rail.htm](http://centralaustincdc.org/transportation/austin_urban_rail.htm)
- Corporation, C. A. (2013). *Austin Urban Rail*. Retrieved from Central Austin CDC:  
<http://centralaustincdc.org/index.htm>
- Cynthia Chen, Jason Chen, James Barry. (2009). Biurnal Pattern of Transit Ridership: a Case Study of the New York City Subway System. *Journal of Transport Geography* , 176-186.
- Filion, P. (2001). Suburban mixed-use centres and urban dispersion: what difference do they make? *Environment and Planning A 33(1)* , 141 – 160 .
- Goldmark, A. (2012, December 09). *Census data show public transit gender gap*. Retrieved from Transportation Nation:  
<http://www.wnyc.org/story/283137-census-data-show-public-transit-gender-gap/>
- Gomez-Ibanez, J. (1996). Big-City Transit Ridership, Defilicts, and Politics: Avoiding Reality in Boston. *Journal of the American Planning Association, Volume 62, Issue 1*, 30-50.
- Javier Gutierrez, Juan Carlos Garcia-Palomares. (2008). Distance-measure impacts on the calculation of transport service areas using GIS. *Environment and Planning B: Planning and Design*, volume 35, 485-503.
- Javier Gutiérrez, Osvaldo Daniel Cardozo, Juan Carlos García-Palomares. (2011). Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. *Journal of Transport Geography*, 1081-1092.
- Jeffrey Brown, Gregory Thompson, Torscha Bhattacharya, Michal Jaroszynski. (2013). Understanding Transit Ridership Demand for the Multidetination, Multimodel Transit Network in Atlanta, Georgia: Lessons for Increasing Rail Transit Choice Ridership while Maintaing Transit Dependent Bus Ridership. *Urban Studies*, 51(5) 938-958.
- John F. Kain, Z. L. (1996). *Econometric analysis of determinants of transit ridership*. Volpe National Transportation Systems Centers, U.S. Department of Transportation.
- Kain, J. F. (1999). Secrets of success: assessing the large increseases in transit

- ridership achieved by Houston and San Diego transit provides. .  
*Transportation Research PartA: Policy and Practice*, 33(7/8): 601-24.
- Kevin Manaugh, L. F.-M.-G. (2010). The effect of neighbourhood characteristics, accessibility, home-work location, and demographic on commuting distances. *Transportation*, 37(4), 627-646.
- Kitamura, R. (1989). A causal analysis of car ownership and transit use. *Transportation*, pp. 155-173.
- Kuby, M., Barranda, A., & Upchurch, C. (2004). Factors influencing light-rail station boardings in the United States. *Transportation Research A* 38 (3), 223-247.
- Norm Marshall, Brain Grady. (2006). Sketch Transit Modeling Based on 2000 Census Data. *Transportation Policy Board, 2006 Annual Meeting CD-ROM*.
- Program, N. C. (2006). *Estimating the impacts of the aging population on transit ridership*. Washington D.C.: Transportation Research Boeard of the National Academics.
- Rober Hannay, M. W. (2006). Factors influencing support for local transportation sales tax measures. *Transportation*, Volumn 34, Issue 1, 17-35.
- Robert Cervero, J. M. (2010). Direct Ridership Model of Bus Rapid Transit in Los Angeles County, California. *Transportation Research Record: Journal of the Transportation Research Board* NO.2104, 1-7.
- Robert Cervero, Kara Kockleman. (1997). Travel Demand and the 3Ds: Density, Diversity, And Design. *Transortation Research, PartD* 2(3), 199-219.
- Robert Cervero, Michael Duncan. (2008). *Residential Self Selection and Rail Commuting: A Nested Logit Analysis*. Berkeley, California: University of California Transportation Center.
- Sale, J. (1976). *Increasing transit ridership: the experience of "seven cities"*. Urban Mass Transportation Administration, U.S. Department of Transportation.
- Samuel Seskin, Cervero, Robert. (1996). *Transit and Urban Form*.
- Sharfuddin J. Syed, A. M. (2000). Factor analysis for the study of determinants of publictransit ridership. *Journal of Public Transportation* , 3(3) 1-17.
- Spillar Robert J., G. Scott Rutherford. (1998). The Effects of Population Density and Income on per Capita Transit Ridership in Western American Cities. *Institute of Transportation Engineers' Compendium of Technical Papers: 60th Annual*

*Meeting*, (pp. 327-331).

Sullivan, M. (2013, January 15). *Gender and Metro Ridership: More Women Ride Than Men, With Some Exceptions*. Retrieved from The Source: <http://thesource.metro.net/2013/01/15/gender-and-metro-ridership-more-women-ride-than-men-with-some-exceptions/>